Disclaimer: The following is non-rigorous and only intended to help clarify ideas for students in the Spring 2014 class of undergrad probability theory.

1 Probability Spaces

Definition 1. A Probability Space is a 3-tuple $(\Omega, \mathcal{F}, P)$ where $\Omega$ is a set called the sample space, $\mathcal{F}$ is a set of subsets of $\Omega$, and $P : \mathcal{F} \to \mathbb{R}_+$ is a probability measure. A set $E \in \mathcal{F}$ is called an event, and an element $\omega \in \Omega$ is called an elementary event.

Remark 2. We think of $\Omega$ as some background space that allows us to encode our information, and $\mathcal{F}$ as the space of things that can happen. Don’t worry about the properties of $\mathcal{F}$. For the purposes of this class, just imagine that it is the power-set (set of all subsets) of $\Omega$.

Not every function $P$ is allowed, only those that are called probability measures

Definition 3. A function $P : \mathcal{F} \to \mathbb{R}_+$ is called a probability measure if

1. $P(\emptyset) = 0, P(\Omega) = 1$.

2. If $\{A_i\}$ is an at most countable collection of disjoint sets, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum P(A_i)$. In particular if two sets are disjoint, $A$ and $B$, then $P(A \cup B) = P(A) + P(B)$

Remark 4. In words Property 1 is says that the probability that nothing happens is 0 and the probability that something happens is 1. Property 2 says if we we have a bunch of events that are mutually exclusive, then the probability that at least one of them happens is the sum of their individual probabilities.

Proposition 5. (Optional) Probability measures satisfy the following monotonicity properties properties:

1. If $A_1 \subset A_2 \subset A_3 \subset \ldots \subset \bigcup A_i$, then $\lim P(A_i) = P(\bigcup A_i)$

2. If $A_1 \supset A_2 \supset \ldots \supset \bigcap A_i$, then $\lim P(A_i) = P(\bigcap A_i)$

Problem 6. Show that $P(A^c) = 1 - P(A)$

Problem 7. A helpful way to translate between set notation and words, Let $A$ and $B$ be to events (i.e. sets) then $A \cap B$ is also some times called “$A$ and $B$” and the set $A \cup B$ is some times called the set “$A$ or $B$”. Explain why this is correct.

Problem 8. Model coin flipping this way using: Heads and Tails, and the numbers 0 and 1. That is, find an appropriate $\Omega$ and $P$. 

Theory Of Probability: Review

Aukosh Jagannath

May 3, 2014
2 Random Variables

Definition 9. A random variable is a function $X : \Omega \rightarrow \mathbb{R}$.

Definition 10. The indicator function of the set $A \in \mathcal{F}$ is the random variable

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

Definition 11. The (cumulative) distribution function of a random variable $X$ is the function $F(x) = P(X \leq x)$.

Problem 12. (Optional) A distribution function has the following properties

1. $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$
2. if $x \leq y$ then $F(x) \leq F(y)$
3. $F$ is continuous from the right.

(Hint: use the monotonicity properties.)

Definition 13. A random variable is discrete if it takes on at most countably many values. Otherwise we’ll call it a continuous random variable.

Definition 14. The (probability) mass function of a discrete random variable is the function $f(x) = P(X = x)$.

Definition 15. The (probability) density function of a continuous random variable is the function $f(x)$ such that

$$\int_a^b f(x) \, dx = P(X \in [a, b])$$

Exercise 16. (Optional) Show that a continuous random variable has a mass function that’s always 0.

(Hint: use the fact that the sum of uncountably many numbers is not defined and property 2 of probability measures.)

Exercise 17. Define a fair coin flip and a biased coin flip with bias $p$ as a random variable on an appropriate sample space. Find its distribution function and plot it. (Note: solve this problem at least 2 different ways. ) Find is probability mass function.

For the following questions, you may go to wikipedia to find the probability density function or mass function.

Exercise 18. Define a gaussian as a random variable. Find a sample space, and a probability measure. Pretend all sets can be generated by the intervals $[a, b]$. (Hint: make your life easy take $\Omega = \mathbb{R}$. Now its up to you to think of the right function and the right probability measure.) Find it’s distribution function and plot it. The PDF is

$$f(x) = \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

Exercise 19. Define an exponential random variable with parameter $\lambda$ using this language, denoted by. The pdf is

$$f(x) = \lambda e^{-\lambda x}$$

Again pretend all sets can be generated by intervals.
Exercise 20. Define a Poisson random variable using this language. The pmf is
\[ f(x) = e^{-\lambda} \frac{\lambda^x}{k!} \]

Exercise 21. If a continuous random variable has a distribution function \( F \), find its probability density function. If it has a density function \( f \), find its distribution function. Give an expression for \( P(X \in (a, b]) \) in terms of the distribution function.

3 Independence

Definition 22. Two events are independent if
\[ P(A \cap B) = P(A)P(B) \]

Two random variables \( X, Y : \Omega \rightarrow \mathbb{R} \) are independent if for any two sets \( A, B \in \mathcal{F} \),
\[ P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \]

Fact 23. If two random variables are independent, then their joint distribution function splits
\[ F_{X,Y}(x, y) = F_X(x)F_Y(y) \]

If they are continuous random variables then their joint density function splits
\[ f_{X,Y}(x, y) = f_X(x)f_Y(y) \]

If they’re discrete, then their joint mass function splits.

4 Expectations

Definition 24. The expectation or expected value or mean of a discrete random variable \( X \) with mass function
\[ \mathbb{E}X = \sum_x xf(x) \]

and the expected value of a continuous random variable with probability density function \( f \) is
\[ \mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx \]

provided the sum and integral are well defined.

Fact 25. Note
1. If \( X \geq 0 \) then \( \mathbb{E}X \geq 0 \)
2. (linearity) if \( a, b \in \mathbb{R} \) then \( \mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y \)
3. \( \mathbb{E}1 = P(A) \)
4. \( \mathbb{E}1 = 1 \)

Definition 26. The \( k \)-th moment of a random variable \( X \) is
\[ \mathbb{E}X^k \]

The variance of a random variable is
\[ \text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 \]
Exercise 27. Show that the variance is the second moment minus the squared mean.

Fact 28. If two random variables are independent then $\mathbb{E}XY = \mathbb{E}X \mathbb{E}Y$ and $Var(X+Y) = Var(X) + Var(Y)$

Exercise 29. Prove this.

5 Generating function and Characteristic function

Definition 30. The generating function for a random variable $X$ is

$$M(t) = \mathbb{E}e^{tX}$$

and the characteristic function is

$$\phi(t) = \mathbb{E}e^{itX}$$

Fact 31. (Lévy) The characteristic function uniquely determines a random variable. For all random variables in our class the random variable is also determined the generating function also determines the random variable.

6 WLLN+CLT+Chebyshev

Theorem 32. (Chebyshev) Let $X$ be a random variable with variance $\sigma^2$. Then

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

Theorem 33. (WLLN) Let $X_i$ be iid with mean $\mu$ and finite second moment. Then

$$\lim P\left(\left|\frac{\sum_{i=1}^{N} X_i}{N} - \mu\right| > \epsilon\right) = 0$$

Theorem 34. (CLT) Let $f$ be a continuous bounded function. Let $X_i$ be as above with variance $\sigma^2$. Then

$$\lim \mathbb{E}f\left(\frac{\sum X_i - n\mu}{\sqrt{n}\sigma}\right) = \mathbb{E}f(Z)$$

where $Z \sim N(0, 1)$

7 Exercises

7.1 IMPORTANT: The Classic random variables

Exercise 35. Compute the first and second moments, the variance, and the generating function of a $Ber(p)$ (coin flip with bias $p$) if you define this as a 0 or 1 valued random variable what does $p$ have to do with the mean?

Exercise 36. Compute the first and second moments, the variance, and the generating function of a $Bin(p, n)$ where it takes values between $[0, n]$. What does this have to do with $Ber(p)$?

Exercise 37. Compute the first and second moments, the variance, and the generating function of a $Geom(p)$.

Exercise 38. Compute the first and second moments, the variance, and the generating function of a $N(\mu, \sigma^2)$. Do the first 3 assuming $\mu = 0$, then calculate the generating function for arbitrary $\mu$ and use that to deduce the answer for the first 3.
Exercise 39. Compute the first and second moments, the variance, and the generating function of an \( \exp(\lambda) \). What is probabilistic interpretation of the parameter?

Exercise 40. Compute the first and second moments, the variance, and the generating function of a Poiss(\( \lambda \)). What is probabilistic interpretation of the parameter?

Exercise 41. (Optional) Using Lévy’s theorem, show that the sum of two gaussians \( N(\mu_1, \sigma_1^2) \) and \( N(\mu_2, \sigma_2^2) \) is a gaussian \( N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \).

Exercise 42. (Optional) Using Lévy’s theorem, show that the sum of two poisson random variables with parameters \( \lambda_1 \) and \( \lambda_2 \) is poisson with parameter \( \lambda_1 + \lambda_2 \).

7.2 CLT+LLN+ Modeling practice

7.2.1 Random Walker

Exercise 43. A simple random walk is defined as follows: Let \( S_0 = 0 \) and then at each time step move +1 or -1 with probability \( 1/2 \). Let \( S_n \) be the position of the random walker at time \( n \). What is the distribution for \( S_n \) in terms of the classic random variables? Can you think of \( S_n \) in terms of sums of random variables? What is the mean at time \( N = 1000 \)? The variance? (less important:) The MGF?

Exercise 44. Roughly speaking, what does the law of large numbers tell you about the position of the walker? That is do you expect the walker to be near \( .1N \) or \( .01N \)? Suppose that the walker has a bias, so that it moves +1 with probability \( p > 1/2 \). Roughly, speaking where do you expect the walker to be at time \( N \)? (If you don’t understand this question after 10 minutes of thinking, then don’t worry about it, it might just be too vaguely worded).

Exercise 45. Using Chebyshev’s inequality, where is the random walker at time \( N \) with probability roughly .95? what is the probability that \( S_N \geq \epsilon N \)? plug in \( \epsilon = 10 \) and \( \epsilon = 100 \).

Exercise 46. In a plot \((x,t)\), draw a graph of roughly what region the central limit theorem tells you to expect the walker to be in at time \( t \). (By roughly I mean with probability \( >.66 \)) In particular, give a formula for the curve \( t = f(x) \) that you should plot.

7.2.2 Planes

Exercise 47. A plane has \( N \) engines. Each engine works with probability \( p \). Let \( E \) be the number of engines that work. What kind of random variable is this? Let \( N = 100 \). What is the mean? variance? MGF?

Exercise 48. Let \( N = 1000 \) and suppose that the plane can land if 500 engines work. Suppose that \( p = 1/2 \). What’s the probability that the plane lands? Suppose that the plane can land if at least 1/3 of the engines work. Should you expect the plane to land? In particular what does the Weak Law of Large Numbers tell you? State the theorem precisely in terms of \( E \).

Exercise 49. Suppose we’re riding this plane together and I’m really panicky. Use Chebyshev and the CLT to convince me that I shouldn’t be worried about the plane landing if it lands provided at least 1/3 work.

Exercise 50. Suppose that you now have to design such a plane. Use the set up from the previous problem. Again using Chebyshev and CLT, about how many engines should you have to expect to land with probability .99.

7.2.3 Dice

Exercise 51. Suppose that the elevators at courant work as follows: when you press the button, an evil gnome repeatedly throws a 10 sided die and waits until he sees a 3 or 4 and then brings the elevator to you immediately. Suppose its a fair die. What kind of random variable is this? what’s your expected waiting time? what’s the variance? what’s the MGF? How long should you wait with probability > .90.
Exercise 52. Suppose I ride the elevator 1000 times a day. Roughly, what is my average waiting time? What theorem guarantees this? What is the 90% confidence interval using: Chebyshev? CLT?

7.3 More Exercises

Exercise 53. Let $A$ be some event. Explain the frequentist interpretation of probability using the (weak) law of large numbers. In particular show that if $X_i$ are iid random variables with finite (second) moment, what is

$$\frac{1}{N} \sum 1_{X_i \in A}$$

for large $N$?

Exercise 54. Suppose that Coca-Cola’s machine makes cans with 330 ml with known standard error of .001 ml. Suppose I come to inspect their factory to see if their machine needs re-calibration. To do this I measure 100 cans of coke and get a sample average of 330.001. Does it?

Exercise 55. Consider the following game: Y have two envelopes with different values of money in them. You do not tell me how much, just that one of them has more money than the other. I get to pick an envelope, look inside, and then decide to switch or stay. Suppose that the way I decide to switch is: pick an exponential random variable with mean 1, if the amount I see is greater than that, stay. Otherwise switch. Show that this is a favorable strategy.

[Hint: Let $X_1$ be the amount in the envelope I’m holding at step 1 of the game and $X_2$ be the amount at step 2. Suppose the amounts are $A$ and $B$ with $A < B$ and assume $P(X_1 = A) = P(X_2 = B) = 1/2$.

1. Calculate the probability $P(X_2 = B|X_1 = A)$
2. Calculate $P(X_2 = B|X_1 = A)$.
3. Use the definitions of conditional expectation to conclude that $P(X_2 = B) > 1/2$.

Exercise 56. The elevators at Courant come so randomly, I model them as a Poisson process. I assume that the expected number of elevators that come to the 13th floor in 5 minutes is 1. What is the distribution of the amount of time I wait for the elevator? Suppose I waited 5 minutes and there was no elevator. Is it more likely that the elevator will come in the next 5 minutes?

Exercise 57. Suppose a monkey throws a die. What’s the expected value of her role? Suppose she rolls $Poiss(1)$ many dice, now what’s the expected value? (Are you surprised by the answer?)